# The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz

Friedrichs, Daniel ; Maurer, Dieter ; Dellwo, Volker

Abstract: In a between-subject perception task, listeners either identified full words or vowels isolated from these words at F0s between 220 and 880Hz. They received two written words as response options (minimal pair with the stimulus vowel in contrastive position). Listeners' sensitivity (A0) was extremely high in both conditions at all F0s, showing that the phonological function of vowels can also be maintained at high F0s. This indicates that vowel sounds may carry strong acoustic cues departing from common formant frequencies at high F0s and that listeners do not rely on consonantal context phenomena for their identification performance.

DOI: https://doi.org/10.1121/1.4922534

# The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz

**Daniel Friedrichs**
*Phonetics Laboratory, Department of Comparative Linguistics, University of Zurich, Plattenstr. 54, CH-8032 Zurich, Switzerland*
*daniel.friedrichs@uzh.ch*

**Dieter Maurer**
*Institute of the Performing Arts and Film, Zurich University of the Arts (ZHdK), Toni-Areal, Pfingstweidstrasse 96, CH-8031 Zurich, Switzerland*
*dieter.maurer@zhdk.ch*

**Volker Dellwo**[a)]
*Phonetics Laboratory, Department of Comparative Linguistics, University of Zurich, Plattenstr. 54, CH-8032 Zurich, Switzerland*
*volker.dellwo@uzh.ch*

**Abstract:** In a between-subject perception task, listeners either identified full words or vowels isolated from these words at $F_0$s between 220 and 880 Hz. They received two written words as response options (minimal pair with the stimulus vowel in contrastive position). Listeners' sensitivity (A′) was extremely high in both conditions at all $F_0$s, showing that the phonological function of vowels can also be maintained at high $F_0$s. This indicates that vowel sounds may carry strong acoustic cues departing from common formant frequencies at high $F_0$s and that listeners do not rely on consonantal context phenomena for their identification performance.



[JH]
**Date Received:** January 21, 2015 **Date Accepted:** June 3, 2015

## 1. Introduction

Vocalic identification in naturally produced vowels at $F_0$s exceeding the $F_1$ they typically reveal in citation-form words has so far mainly been a concern of singing research, particularly in Western classical singing ("legitimate style," henceforth: legit). By now there is a large body of evidence indicating that the identifiability of vowels decreases with increasing $F_0$. Identification of single vowels has been shown to be compromised when $F_0$ significantly exceeds $F_1$; also referred to as "oversinging" (Smith and Wolfe, 2009). Early evidence for this position goes back to self-experiments by von Helmholtz (1885:110) who found that the vowel /u/ loses its typical timbre from the musical note $F_3$ (≈175 Hz) upwards and shifts toward /o/. Howie and Delattre (1962) showed in an experiment with nine isolated sung vowels that listeners' identification performance decreased when $F_0$ exceeded $F_1$. Hollien *et al.* (2000) showed for /i/, /a/, and /u/ that vowel category perception shifted mainly to the one with the next higher $F_1$ as $F_0$ increases (i.e., /i/ shifted to /ɪ/ then /ɛ/ then /a/ when $F_0$ exceeded $F_1$ for /i/, and /u/ shifted to /ʊ/, /ɔ/, /ʌ/, /a/, respectively). Other studies have been more precise in identifying an absolute frequency at which the identification performance of listeners decreases. Sundberg (2012) provided evidence that this point corresponds to the musical note *C5* (≈523 Hz). Above this frequency, identification is heavily biased toward open vowels like /a/; from around 700 Hz it arrives at chance performance.

In legit, the communicative aim of producing intelligible utterances is typically in competition—and possibly secondary—to the aim of producing esthetical, sonorant, and powerful vocalizations. Legit singers, for example, adopt their resonance frequencies with the aim of enhancing their vocal power and homogeneity of timber, albeit at the expense of intelligibility (Joliveau *et al.*, 2004). It is probably the result of this subordinate relevance of the communicative function of vowels in legit singing that vowel identification has primarily been studied in isolated vowels in the singing literature. From a linguistic point of view, however, the interest in vowels is typically in the functions they fulfill in speech communication like their phonological function in linguistic contrastive position (e.g., /ɛ/ and /ɪ/ may distinguish between the words *desk* and *disc*).

[a)]Author to whom correspondence should be addressed.

 

Given the evidence for impoverished vowel identification performance of naturally produced vowels at high $F_0$s above $C5$ from the singing literature, it seems conceivable that the phonological function of vowels in minimal pairs should also decrease with a substantial increase in $F_0$ above this frequency. In the phonetic literature this question has so far not received much attention. Key studies on vocalic variability (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Pätzold and Simpson, 1997) were primarily concerned with vowels at relatively low $F_0$s (i.e., substantially below $F_1$ in citation-form words). This is also in line with observations that machine measurements of formants based on standard procedures (e.g., Linear Prediction Analysis) are highly problematic when $F_0 > C5$. It seems that it is implicitly taken for granted by phoneticians that the phonological function of vowels at high $F_0$s should thus be poor. And this assumption seems justified as the probably strongest cues to vowel category identification—formant frequencies (in terms of determinable spectral maxima)—are poor when $F_0$ increases significantly above $C5$.

Evidence exists which indicates that the consonantal environment of vowels at high $F_0$s in real words enhances vowel identification. Smith and Scott (1980) reported higher identification rates for the front vowels /i/, /I/, /ɛ/, /æ/ at $F_0$s up to about 1100 Hz when they were produced in word consonant-vowel-consonant (CVC) context (/b/-V-/d/ resulting in *bead*, *bid*, *bed*, and *bad*) compared to the same vowels produced in isolation. One might assume that such results are driven by formant-transition phenomena between consonants and vowel (Strange *et al.*, 1976), however, their impact on vowel identification has been strongly put into question (Diehl *et al.*, 1981). It seems more likely that co-articulatory phenomena can explain the effect in Smith and Scott, as the vocal tract configuration of a vowel is to a large degree in position during the surrounding consonants. This is particularly audible, when one of the consonants is a voiceless fricative, characterized by a broadband noise source and ideally produced toward the rear end of the vocal tract (e.g., /heed/ and /hood/). In this case listeners can likely profit from the acoustic characteristics of the noise source shaped by the co-articulated vocal tract resonances of the vowel.

It also seems plausible that the poor identifiability of vowels at $F_0$s higher than $C5$ is to a considerable degree the result of legit singing. This has already been suggested by Sundberg (2012) in particular, with reference to Smith and Scott (1980) who showed that in legit style, vowel intelligibility was poorer than in a condition in which singers raised their larynx and thus adapted their resonances to the increased $F_0$. Such evidence for a better identifiability of vowels at high $F_0$s in a non-legit style was provided by Maurer *et al.* (2014) for a female singer of Cantonese opera. Listeners identification performance was drastically better than chance for 4 of her vowels (/i/, /a/, /ɔ/, /u/) up to an $F_0$ of 860 Hz. Because of the strong focus on voice esthetics in the singing literature, it remains unclear to what degree the phonological function is maintained at high $F_0$s when a singer focuses on intelligibility rather than esthetics (i.e., when the singer does not sing in a specific singing style).

Here we asked a trained female singer to produce minimal pairs including all long vowels of her native language (German) at varying $F_0$ levels between 220 and 880 Hz focusing on the intelligibility of speech and, if necessary, ignoring esthetic qualities of her singing style. We extracted the steady state vocalic part (always 250 ms) of the word productions, resulting in two experimental conditions, words and isolated vowels. The fact that we made the singer produce the two words of each minimal pair in sequence, inevitably made her focus on the phonologically contrastive nature of the vowels during the production. In a between-subject design perception task, two groups of German native listeners identified the words extracted from the minimal pair productions being either presented as a full word stimulus (condition 1) or an isolated vowel (condition 2). We extracted the words from pairs in which the difference in $F_1$ is expected to play a crucial role in the distinction of the vowels. This is true in particular, in minimal pairs contrasted by the front vowels /i/, /e/, /ø/, /y/, /ɛ/, /a/ (15 possible pairs) and by the back vowels /u/, /o/ together with /a/ (3 possible pairs) in which between-category variability of $F_2$ is comparatively low but high for $F_1$. We tested to what degree listeners' ability to identify the correct word of a minimal pair decreased with increasing $F_0$ for all vowel pairs. To avoid having varying numbers of response options and to test the words from the original pair productions, we provided listeners with binary response options (two words of the minimal pair). Should it hold that vowels with an $F_0 > C5$ lack acoustic category information then we would expect that: (i) For high-back vowels with low $F_1$ and low $F_2$, word identification performance should be poorest. (ii) Vowels in which $F_0$ exceeds $F_1$ should more often be perceived as /a/-like, so for minimal pairs in which a contrast is built with the vowel /a/ listeners' identification performance should drop with

higher $F_0$ and listeners should be biased in their perception toward /a/. (iii) Should listeners rely on consonantal environment effects (co-articulation or formant-transitions), it should be expected that identification performance drops drastically when such information is removed in vowels extracted from the carrier word (condition 2).

## 2. Methods

### *2.1 Subjects*

Forty native German listeners without reported hearing impairments [20 male, 20 female; mean age = 26.78, standard deviation (s.d.) = 7.43], all students at the University of Zurich, participated in the experiment. Listeners were randomly divided into two groups ($N = 20$ per group; one group per condition [word and isolated vowel]; gender balanced across groups; mean age group 1: 29.75, s.d. = 8.73, group 2: 23.8, s.d. = 4.29).

### *2.2 Stimuli and apparatus*

One female Musical Theatre singer (age 33; Swiss German native speaker, with excellent and trained pronunciation of Standard German) was recorded with a cardioid condenser microphone (Sennheiser MKH 40 P48 with pop shield, Wedemark-Wennebostel, Germany) on a PC via an audio interface (Fireface UCX, RME, Halmhausen, Germany) in a noise-controlled room at the University of Zurich. The singer was recorded in standing position; a drawn position reference on the floor helped the singer to keep a constant distance of about 30 cm to the microphone. The singer was selected based on her extended vocal range and a high skill of maintaining vowel quality at high $F_0$s. The singer produced 18 German minimal pairs with a vocalic contrast in word mid position. All words were disyllabic and the contrasted vowels were part of the first syllable. Each contrastive vowel was in a CVC syllable. Mean duration of the vowels was 0.68 s (range: 0.58–1.11 s). Two sets of vowel contrasts were built, one with front vowels (/i:/, /y:/, /e:/, /ø:/, /ɛ:/, /a:/) and one with back vowels together with /a:/ (/u:/, /o:/, /a:/). All vowels were contrasted with each other within the two different sets:

- Fifteen front vowel pairs: Biene-Bühne (/i:/-/y:/), siegen-Segen (/i:/-/e:/), biegen-Bögen (/i:/-/ø:/), schielen-schälen (/i:/-/ɛ:/), siegen-sagen (/i:/-/a:/), lügen-legen (/y:/-/e:/), rühren-Röhren (/y:/-/ø:/), schürfen-schärfen (/y:/-/ɛ:/), Sühne-Sahne (/y:/-/a:/), Lehne-Löhne (/e:/-/ø:/), legen-lägen (/e:/-/ɛ:/), Segen-sagen (/e:/-/a:/), töte-täte (/ø:/-/ɛ:/), Söhne-Sahne (/ø:/-/a:/), schälen-Schalen (/ɛ:/-/a:/).
- Three back vowel pairs (including /a:/): Buden-Boden (/u:/-/o:/), Buden-baden (/u:/-/a:/), Boden-baden (/o:/-/a:/)

The word pairs were recorded in two runs in AB and BA order. The singer was instructed to produce the minimal pairs as intelligible as possible. The word pair (AB or BA) that appeared to have the more perceptually salient vowel contrast to an investigator (second author) was chosen for the investigation. Each word pair was recorded at nine $F_0$ levels (220, 440, 587, 659, 698, 740, 784, 831, 880 Hz) resulting in 162 minimal pairs (9 frequencies × 18 vowel contrasts). The lowest $F_0$ level corresponded to the average $F_0$ in citation-form words (Hillenbrand *et al.*, 1995) and the entire frequency range of $F_0$ produced was the range of the average $F_1$ for German vowels produced by women (Pätzold and Simpson, 1997). The respective piano notes were presented as reference sounds to the singer via loudspeaker immediately preceding the production. $F_0$ of the sound produced was measured in Praat (Boersma and Weenink, 2015) in the extracted vocalic parts. A maximum deviation from the reference $F_0$ of 2.5% was found. Each of the two words from the chosen word pair recordings was extracted to serve as a stimulus in the word condition. For the isolated vowel condition, the steady state vowel centers were extracted with a duration of 250 ms (±125 ms from the vowel mid point). At on- and offset the sounds were faded over 50 ms by amplitude modulating the waveform with half a period of a cosine function [fade-in: $(1 - \cos(x))/2$; fade-out: $(1 + \cos(x))/2$]. Each stimulus was normalized for intensity (0 dB difference between stimuli); the overall output level was chosen by listeners individually.

### *2.3 Procedure*

Two word identification tests were carried out (one for each condition) in a small and noise controlled room using closed dynamic headphones (Beyerdynamic DT 770 Pro, 250 Ω). In test 1, listeners were presented each word from each minimal pair ($N = 324$; 9 frequencies × 18 minimal pairs × 2 words) and saw a screen that contained 2 buttons (horizontally arranged) labeled with the words of the minimal pair (position—left/right—was chosen randomly for each response option set). Above the response buttons the sentence *Welches Wort hörst Du?* (English: *Which word do you hear?*) could be

read. Listener's task was thus to identify the word presented from the two response options (minimal pair) provided. Mm. 1 contains an example of a word stimulus and Mm. 2 the respective isolated vowel stimulus derived from this word.

Mm. 1. Word stimulus "Buden" at 880 Hz; response options = "baden" and "Buden". This is a file of type "wav" (118 Kb).

Mm. 2. Isolated vowel stimulus /u:/ at 880 Hz extracted from the word Buden in Mm. 1; response options = "baden" and "Buden". This is a file of type "wav" (21 Kb).

After listeners made their choice they would hear the next stimulus automatically with a delay of 1 s. Listeners could not repeat a stimulus. Test 2 was identical to test 1 with the exception that an isolated vowel instead of a word was presented for identification. Above the response buttons, listeners could read the sentence *Aus welchem Wort stammt der Vokal?* (English: *From which word did this vowel derive?*). In test 2, listeners were explained that the presented vowel only referred to the contrasting vowel in the first syllable of the disyllabic word.

### *2.4 Data analysis*

Listeners' identification performance was calculated with the bias free non-parametric sensitivity measure A′ from Signal Detection Theory (Stanislaw and Todorov, 1999) with Praat scripts written by V. Dellwo according to formulas in Pallier (2002). One of the response options was arbitrarily assigned to the signal (signal vowel), the other to the noise (noise vowel). A "hit" was thus *signal vowel presented and responded*, a "miss" was *signal vowel presented but not responded*, a "false alarm" was *noise vowel presented but not responded*, a "correct rejection" was *noise vowel presented and responded.* A′ ranges between 0 and 1 with 0.5 being chance performance and 1 maximum performance. Values below 0.5 indicate response confusion. Listeners' response bias (i.e., a bias toward the vowel /a:/; see Sec. 1) was measured by $B''_D$ (Pallier, 2002). $B''_D$ ranges from −1 (maximum noise bias) to +1 (maximum signal bias). As each vowel was presented only once per listener, we pooled over listeners ($N = 20$) to calculate A′ for each vowel pair at each $F_0$ level and signal condition ($N = 40$; for example, the pair /i:/ vs /e:/ was presented 20 times for /i:/ and 20 times for /e:/). So each A′ value was calculated based on 40 responses by 20 listeners to a vowel pair.

## 3. Results

Figure 1 shows the distributions of A′ at each $F_0$ for the word and isolated vowel conditions of all minimal pairs; Fig. 2 shows the A′ for word and isolated vowel conditions for each of the 18 minimal pairs separately. A′ values for all investigated $F_0$ levels (i.e., 220–880 Hz) are high above chance level for both the word and isolated vowel conditions. For the word condition performance is at ceiling throughout all $F_0$ levels. For the isolated vowel condition the interquartile range is roughly between A′ 0.9 and 1 at higher $F_0$ levels. Two one-sample *t*-tests (one per condition; alpha = 0.01) testing the mean of the distribution against A′ chance level (0.5) show that the effect was highly significant in both cases (words: $t^{17} = 83.43$, $p < 0.001$; isolated vowels: $t^{17} = 29.23$, $p < 0.001$). The poorer performance for isolated vowels in comparison to words was highly significant (Welch two-sample *t*-test: $t[222.75] = 7.32$, $p < 0.001$). To test that this effect could be replicated for individual $F_0$ levels we carried out 18 one-sample *t*-tests, one for each $F_0$ level (Bonferroni correction = 0.05/18 = 0.0028). *T* for 17 degrees of freedom ranged from 28.14 to 534.62. Each effect was highly significant ($p < 0.00028$).

To test the variation of A′ between $F_0$ levels we carried out a 9*2 two-factor analysis of variance (ANOVA) ($F_0$* condition). Results revealed a highly significant interaction ($F^{8,306} = 2.92$, $p < 0.005$) which was why we proceeded to calculate simple effects for each factor. Simple effects for $F_0$ were studied by two one-factor ANOVAs (one for each condition). The effect for the word condition was not significant ($F^{8,153} = 1.01$, $p = 0.39$) and highly significant for the isolated vowel condition ($F^{8,153} = 5.14$, $p < 0.001$). This means that listeners had equally high performance in the word condition at all $F_0$ levels and that performance decreased significantly with $F_0$ in the isolated vowel condition. Simple effects for condition were tested by 9 two-sample *t*-tests (Welch) with a Bonferroni corrected alpha level of 0.0055 (0.05/9 $F_0$ levels). A significant effect could be obtained for $F_0$ level 4 (659 Hz) ($t[22.94] = 3.25$, $p < 0.005$) and a highly significant effect for level 9 (880 Hz) ($t[22.46] = 4.3$, $p < 0.0005$). It was surprising to obtain a significant effect at level 4 but not at the next higher levels (until level 9).

Listener bias calculation toward /a:/ ($B''_D$) is not meaningful when A′ is high as it is only based on a small number of misses/false alarms (Stanislaw and Todorov, 1999). For this reason, we calculated $B''_D$ only in case of the vowel pair /a:/-/ɛ:/ under the isolated vowel condition for $F_0$ of 831 and 880 Hz where A′ values dropped to 0.81 and 0.75, respectively

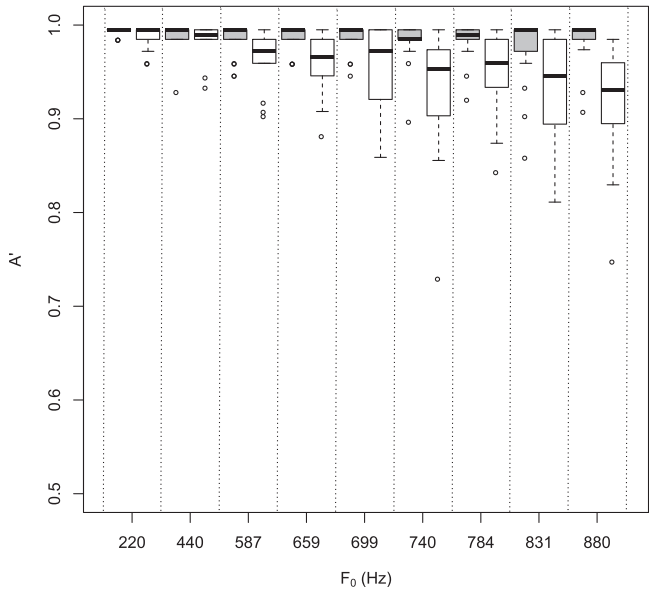


Fig. 1. Box plots showing the distributions of A′ (*y* axis) for all vowel pairs that were tested at nine $F_0$ levels (*x* axis). condition 1, words: white; condition 2, isolated vowels: gray. A′ reaches from 0.5 (chance) to 1 (maximum performance).

(Fig. 2). We received $B''_D$ values of 0.8 and 0.89, respectively, indicating a strong signal bias (i.e., /a:/). This is small evidence for the hypothesis that under severe listening conditions (isolated vowels), listeners are biased in their perception of /ɛ:/ toward /a:/ vowels at high $F_0$s. However, this does not hold true for all other vowel contrasts tested that included /a:/ because the general performance for these vowel pairs was too high. For the high-vowels together with /a:/ (/a:/-/i:/ and /a:/-/u:/), where the strongest decrease in performance should be expected because $F_0$ exceeds $F_1$ drastically, the word identification performance was at ceiling level in both the word and the isolated vowel conditions.

Rare cases of higher A′ for vowels tested in isolation compared to vowels tested in words could also be observed. This was true for /ø:/-/a:/ at $F_0 = 220$ Hz, /ɛ:/-/a:/ at $F_0 = 440$ and $F_0 = 587$ Hz, /o:/-/a:/ at $F_0 = 659$, $F_0 = 699$, and $F_0 = 740$ Hz, /i:/-/ɛ:/ at $F_0 = 699$ Hz, /y:/-/ø:/ at $F_0 = 440$ Hz, /y:/-/ɛ:/ at $F_0 = 440$ Hz, /e:/-/ɛ:/ at $F_0 = 831$ Hz, and /ɛ:/-/ø:/ at $F_0 = 587$ and $F_0 = 831$ Hz (Fig. 2). As these cases occurred non-systematically

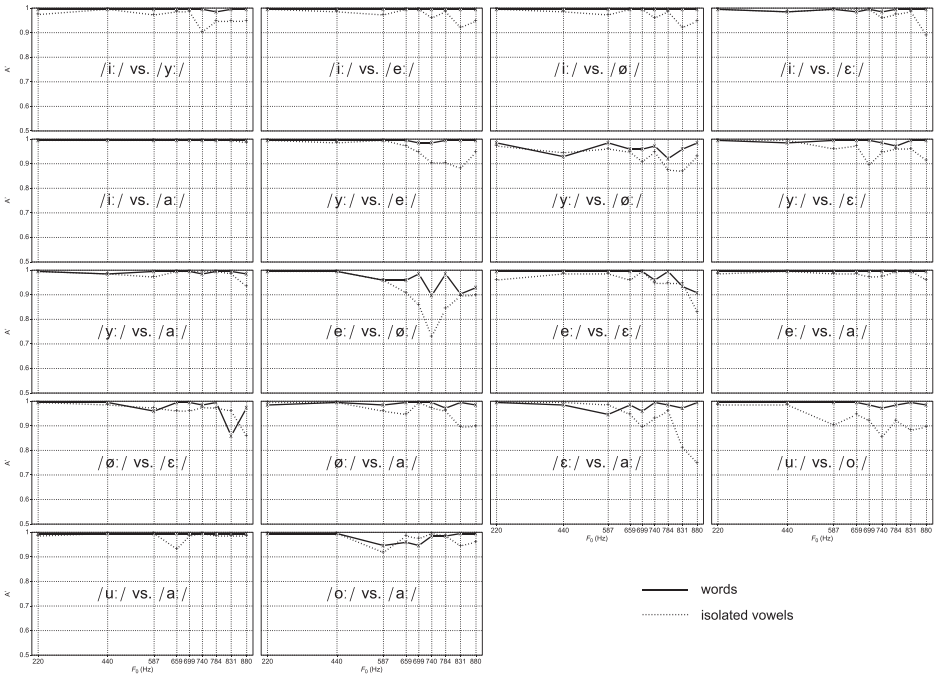


Fig. 2. A′ (*y* axis) for words (solid lines) and isolated vowels (dotted lines) for each of the minimal pair contrasts at the nine investigated $F_0$ levels (*x* axis). A′ reaches from 0.5 (chance level) to 1 (maximum performance).

it seems likely that this was random variability or production variability in the data. It is unlikely that the speaker produced all vowel contrast equally well in each case.

## 4. Discussion

Results revealed that the phonological function of vowels can be surprisingly well maintained up to an $F_0$ of at least 880 Hz. Even though an effect of signal condition (word vs isolated vowels) was obtained, it must be concluded that the performance was extremely high under both conditions. The fact that the identification performance based on isolated vowels was only little below the performance of full word identification and always significantly above chance, is support for the view that the isolated steady state part of the vowel contains sufficient vowel category information even at $F_0 = 880$ Hz. It means that listeners do not rely on possible co-articulatory or formant-transition information in the surrounding consonants for their identification. What is the reason for this high identification performance in the isolated condition? It is possible that vowels produced in a linguistically meaningful environment contain clearer acoustic information to their category, in particular, when produced under severe conditions like at an $F_0$ of 880 Hz. Isolated vowels which were produced in isolation by a speaker (Smith and Scott, 1980) resulted in lower identification results compared their context. It might also be the reason why Deme (2014) found no increase in performance of vowels in nonsense context environment in comparison to isolated vowels.

Listeners' ability to identify a word correctly in the word stimulus condition (condition 1) did not significantly decrease with increasing $F_0$ up to 880 Hz for all vowel pairs tested. This was also true for the high back vowels for which we expected a strong decrease in performance. Therefore, we conclude that an increasing spectral under-sampling, which should inevitably lead to poorer vowel identification accuracy because of the sparser distribution of the harmonics, does not generally lead to a deterioration of the phonological function of vowels. In the case of isolated vowels, performance deteriorated significantly within a range of $F_0$ from 220 toward 880 Hz. This might be weak evidence for a decrease in performance with the loss of consonantal context at vowels with $F_0 > C5$. It is also possible that the artificially generated fading at on- and offset in extracted vowels creates artifacts which contribute to this effect.

What role did $F_1$ and $F_2$ play for our results? It is unlikely that $F_1$ played a crucial role for vowel identification within a vowel pair concerning sounds at very different levels of $F_0$. Words with high vowels containing maximally low $F_1$ and back vowels containing additionally maximally low $F_2$ (/i:/, /y:/, /e:/, /u:/, /ø:/ and /o:/) could typically be identified at ceiling level across all $F_0$ levels. We thus provided an example in which the phonological function of vowels is perfectly maintained when $F_0$ substantially exceeds $F_1$. Concerning $F_2$, the pairs /u:/-/o:/ and /y:/-/ø:/ in long German vowels are strongly under-sampled by $H1$ and $H2$ when $F_0 = 880$ Hz (see Sec. 1). In the case of /y:/-/ø:/ the average $F_2$ frequencies in German are very close (1667 and 1646 Hz, respectively; Pätzold and Simpson, 1997). With an $H2$ at 1760 Hz it seems highly unlikely that $F_2$ was realized in a way in which it could contain subtle cues to vowel category in adjacent high back vowels. It thus seems unlikely that $F_2$ aided listeners in the word identification task in such cases. It is possible, however, that the position of vocal tract resonances between the harmonics influences the relative amplitude of higher harmonics which may in return contain cues to vocalic category. To estimate the frequency of a vocal tract resonance by the relative harmonic amplitudes it is necessary for the listener to have experience with the spectrum of the vocal source. On the one hand it seems feasible that such knowledge was built up over the course of the experiment; on the other hand, we did not find any evidence that listeners performed less well for stimuli at 880 Hz when they incidentally occurred at the very beginning of the randomized stimulus set presentation. Future research will need to test whether listener's identification performance at $F_0 = 880$ Hz improves with knowledge of a speaker's voice.

Listener bias toward /a:/ could typically not be tested in the minimal pairs containing this vowel as listeners' sensitivity was too high. The two cases, however, in which the performance allowed measuring listener bias revealed that a bias toward /a:/ was present. Under more severe listening conditions or with more inexperienced speakers it seems conceivable that such an effect might occur more often.

Given the diverging results from previous studies, it is possible that individual speakers have a high impact on the results. Our speaker was a professional singer in Musical Theater style singing (i.e., non-legit) and is thus probably better suited to depart from legit's esthetic resonance requirements. It thus seems feasible that our speaker was particularly well able to produce the vocalic contrastive information at high $F_0$ levels due to extended vocal range, articulation, and professional training. Our

example, however, proves that it is generally possible for speakers to produce vowels containing sufficient contrastive information at high $F_0$s for reliable identification based on word presentations or isolated vowels. This finding is surprising, also for an individual speaker. It stands in contrast to the widely held view that cues to vowel category at $F_0$s exceeding $F_1$ are technically impossible to produce. To generalize our findings, however, it will be important to study vowel recognition at high $F_0$s with more speakers and possibly a larger variety of response options.

The finding now poses the question about which acoustic cues are responsible for the high word identification performance. Given that our speaker was able to produce contrasts between adjacent high vowel pairs (front as well as back pairs) which should be most affected by high $F_0$s, it puts doubt on the widely held view that formant frequencies were the dominant cues in the word identification tasks. It is possible that other cues such as vowel inherent spectral change (Nearey and Assmann, 1986) explain the performance. The steady state parts in our vowels, however, did not show typical spectral dynamic phenomena of continuous speech or isolated vowel productions. It thus seems questionable to what degree such phenomena might really explain listener identification performance in our vowels. Whichever cues future studies will reveal to be responsible for the result, it is possible that the cues to vowel identity at these high $F_0$s might change our understanding of such cues at $F_0$s typical for conversational speech (Maurer *et al.*, 2000) and might thus contribute highly to our general understanding of human vowel perception.

## Acknowledgments

We thank the professional singer, Heidy Suter, for producing the vowels for this study. This work was supported by the Swiss National Science Foundation (SNSF), Grant No. 100016_143943/1, and the Forschungskredit of the University of Zurich, Grant No. FK-14-062.

## References and links

Boersma, P., and Weenink, D. (**2015**). "Praat: Doing phonetics by computer [Computer program]," Version 5.4.08, retrieved March 30, 2015 from http://www.praat.org/ (Last viewed March 30, 2015).

Deme, A. (**2014**). "Intelligibility of sung vowels: The effect of consonantal context and the onset of voicing," J. Voice **28**, 523.e19–523.e25.

Diehl, R. L., McCusker, S. B., and Chapman, L. S. (**1981**). "Perceiving vowels in isolation and in consonantal context," J. Acoust. Soc. Am. **69**(1), 239–248.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hollien, H., Mendes-Schwartz, A. P., and Nielsen, K. (**2000**). "Perceptual confusions of high-pitched sung vowels," J. Voice **14**(2), 287–298.

Howie, J., and Delattre, P. (**1962**). "An experimental study of the effect of pitch on the intelligibility of vowels," NATS Bull. **18**, 6–9.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.

Maurer, D., D'Heureuse, C., and Landis, T. (**2000**). "Formant pattern ambiguity of vowel sounds," Int. J. Neurosci. **100**, 39–76.

Maurer, D., Mok, P., Friedrichs, D., and Dellwo, V. (**2014**). "Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer," in *15th Annual Conference of International Speech Communication Association*, pp. 2132–2133.

Nearey, T., and Assmann, P. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Pallier, C. (**2002**). "Computing discriminability and bias with the R software," URL: http://www.pallier.org/ressources/aprime/aprime (Last viewed June 12, 2015).

Pätzold, M., and Simpson, A. (**1997**). "Acoustic analysis of German vowels in the Kiel Corpus of read speech," Arbeitsberichte des Instituts für Phonetik und Digit. Sprachverarbeitung Univ. Kiel **32**, 215–247.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Smith, J., and Wolfe, J. (**2009**). "Vowel-pitch matching in Wagner's operas: Implications for intelligibility and ease of singing," J. Acoust. Soc. Am. **125**, EL196–EL201.

Smith, L. A., and Scott, B. L. (**1980**). "Increasing the intelligibility of sung vowels," J. Acoust. Soc. Am. **67**, 1795–1797.

Stanislaw, H., and Todorov, N. (**1999**). "Calculation of signal detection theory measures," Behav. Res. Methods, Instrum., Comput. **31**, 137–149.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.

Sundberg, J. (**2012**). "Perception of singing," in *Psychology of Music*, 3rd ed., edited by D. Deutsch (Academic Press, London), pp. 69–106.

von Helmholtz, H. (**1885**). *On the Sensation of Tone* (Dover, New York), republication 1954, 2nd ed. of the Ellis translation from 1885.